Assessment of historical figures using BabelNet^{TM}

Maciej Kowalski
National Information
Processing Institute
al. Niepodleglosci 188B
00-608 Warsaw
`mkowalski@opi.org.pl`

Szymon Roziewski
National Information
Processing Institute
al. Niepodleglosci 188B
00-608 Warsaw
`sroziewski@opi.org.pl`

Abstract

Sentiment analysis is an important part of natural language processing, which aims to determine a sentiment (or emotion) carried within analyzed text. These analysis can be based on simple human-tagged dictionaries, or more sophisticated machine learning methods. Sentiment analysis is widely used in marketing, analyzing social media, evaluating questionnaires. The purpose of this paper is to describe how to build a simple sentiment analysis system based on *BabelNet* and sentiment dictionaries (both sentiwordnet and lexicon), which can be used to assess historical figures - politicians, scientists, sport figures, actors. Key part of the proposed solution is to query *BabelNet* for synsets surrounding an entity (historical figure). Query can work recurrently, going deeper into synsets and producing a synset tree which surrounds an entity. In *BabelNet*, each synset carries short gloss - definition for given synset. By adding sentiment analysis on those glosses, one can calculate sentiment for a tree of synsets and ultimately assess the person in question.

1 Introduction

Based on the fact that one of the sources for building semantic network is Wikipedia, the *BabelNet* holds the ability not only to query for objects, but also to store information about people, who played, or are still playing an important role in history. It can be tested by querying for example "Ernest Hemingway" using *BabelNet* web site[1]. The response for such query would be typically a set of named entities.
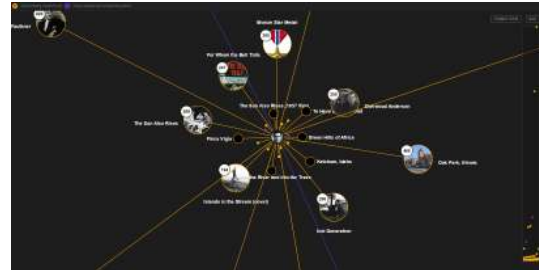
Once the entity is selected, user can explore the semantic network, where the closest entities would appear, giving a semantic graph. By visiting each graph node and gathering all the glosses, user can build a small corpus of text surrounding that entity.

2 Sentiment Analysis

After collecting all the glosses, they can be cross-checked with simple sentiment analysis approach - sentiment dictionaries. In order to proceed with this step, two dictionaries are used:

- *lexicon* dictionary [2]

- *sentiwordnet* [3]

First dictionary has been divided into two parts: positive and negative words. The second dictionary holds a list of words all of which are described by two real numbers from 0 to 1 one referring to positive impact, the other one referring to negative impact.

In order to correctly calculate the sentiment on words in extracted glosses, one must first get the basic forms of those words. In order to do that, the *morfologik* [4] library has been used.

During the evaluation, each word of a gloss has been taken into account and overall positive and negative impacts have been calculated, by summing up all word sentiment results.

---

[1] http://babelnet.org/

[2] http://www.cs.uic.edu/ liub/
[3] http://sentiwordnet.princeton.edu/
[4] http://morfologik.blogspot.com

Table 1: Historical figures with impact scores

historical figure	positive / negative impact
Albert Einstein	0.523 / 0.477
Isaac Newton	0.700 / 0.300
Al Capone	0.304 / 0.696
Marilyn Monroe	0.602 / 0.398
Joseph Stalin	0.448 / 0.552
Billy the Kid	0.601 / 0.399
Charles Manson	0.371 / 0.629

## 3 Implementation and results

The described process has been automated by using Java along with *BabelNet* API for querying the semantic network. To select the synset for given query, following parameters were used:

- Language = EN

- BabelPOS = NOUN

- BabelSenseSource = WIKIDATA

For testing purposes, a few historical figures has been selected and assessed with given algorithm. The scores are presented in *Table 1*.

For given set of names it can be said that following approach can be used to assess historical figures. People like Albert Einstein, Isaac Newton, Marilyn Monroe can be considered as ones having positive impact, on the other hand individuals like Al Capone, Joseph Stalin and Charles Manson are associated with crimes and negative impact on human history. An interesting case comes wit Billy the Kid, who was the notorious gunslinger and should be associated with criminal activities. As it turned out, the entity selected by *BabelNet* referred to movie carrying the same title Billy the Kid.
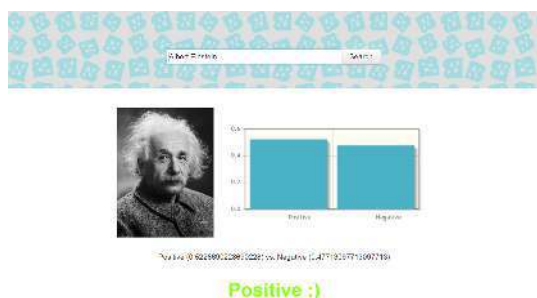
Described algorithm has been also encapsulated into one-page web application [5] based on Java EE

---

[5]https://bhs.opi.org.pl/bhs-web/

technology. The source for this application containing *BabelNet* querying methods and sentiment calculations is available on GitHub [6]

4 Conclusions and further work

It has been shown that the problem of assessing historical figures can be dealt with by using *BabelNet* and simple sentiment analysis system based only on dictionaries. Although there were no gold-standard sets to compare with, by checking dozens of most popular historical figures, we can assume that such method can be developed further. One of the key ingredient to this method would be to use disambiguating mechanism in order to select the exact synset carrying glosses related to historical figure in question.

References

Bing Liu. 2010. *Sentiment Analysis and Subjectivity.*, An chapter in Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkhya and F. J. Damerau).

R. Navigli and S. Ponzetto. 2012. *BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network.*, Artificial Intelligence, 193, Elsevier, 2012, pp. 217-250

S. Baccianella and A. Esuli and F. Sebastiani 2010. *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.*, Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 2010

A. Esuli and F. Sebastiani 2006. *SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining.*, In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06), 2006, pp 417-422

---

[6]https://github.com/dotjabber/babelnet-historical-sentiment